PEPFAR Spatial Data Life Cycle

March 2023

U.S. DEPARTMENT OF STATE

Contents

1. Purpose and scope

The Spatial Data Lifecycle is a data stream-specific complement to broader Data Governance policy and draws upon relevant directives and existing policy documents and guidance from the Office of Management and Budget (OMB) and within the U.S. Department of State and PEPFAR implementing agencies.[1] Spatial data are the latitude and longitude we use to represent health facilities in maps and geographic analysis. They also include the polygons that demarcate administrative units in Panorama and DATIM. Spatial data use supports PEPFAR's work to achieve sustained control of the HIV/AIDS epidemic. This document provides information on how PEPFAR data stewards create, receive, and manage spatial data. This spatial data lifecycle enhances the quality, consistency, availability, and utility of these data for programming and policy decisions.

2. Varieties of spatial data in PEPFAR

Most spatial data is in a raster or vector format. Rasters are grids where each cell contains a value. In the PEPFAR program we use this data format when working with satellite imagery or representations of population distribution. Vector data can represent points, lines, and polygons for mapping roads, administrative units, cities, health facilities, and other features. PEPFAR uses a wide array of spatial data in its analyses but manages only two types: points and polygons. Points capture the locations of structures, such as PEPFAR-supported health facilities, labs, and medical stores. Polygons demarcate areas used to represent national boundaries and internal administrative divisions such as provinces or health districts.

3. Points

DATIM is the system of record for the geographic coordinates of PEPFAR-supported facility-level sites. With two exceptions, facility-level sites that PEPFAR supports should have associated geographic coordinates in DATIM. First, DATIM may include facility-level sites that PEPFAR does not support. The management of latitude and longitude for non-PEPFAR-supported sites is useful for some analyses but is not required. Second, PEPFAR does not capture or maintain geographic coordinates for military sites (e.g. bases, barracks, or other military locations). All other facility records should include accurate latitude and longitude in DATIM.

Facility latitude and longitude in DATIM may originate in partner government systems and may not be U.S. Government data. In some cases, partner governments categorize health facility spatial data as open data while other countries could have more restrictive policies. Therefore, health facility latitude and longitude are restricted data and not shareable outside of USG without partner government clearance by default. Country teams should identify the conditions under which PEPFAR can distribute spatial data in our effort to support partner government responses to the local epidemic.

[1] PEPFAR Data Governance Guidance, https://datim.zendesk.com/hc/en-us/articles/360061233211-PEPFAR-Data-Governance-Guidance

PEPFAR encourages the publication of facility lists with geographic coordinates as open data. With very few exceptions, facility name, type, ownership, administrative area, and geographic coordinates are not inherently sensitive information. Furthermore, the coordination function of a high quality, authoritative facility list derives from people's ability to access and use the list. Broad accessibility of a reference list supports consistency across systems that rely on a list and reduces duplication of effort that attends the maintenance of multiple lists.

3.1 Points: Roles and responsibilities

3.1.1 DATIM Site Administrators

DATIM Site Administrators are responsible for maintaining geographic coordinates in DATIM and should work with partner government counterparts, implementing partners, and PEPFAR team members to prioritize the collection and improvement of coordinates. DATIM Site Administrators should conduct basic quality checks on spatial data before entering them into the system (see below).

3.1.2 Program Results and Impact Monitoring for Epidemic Control (PRIME)

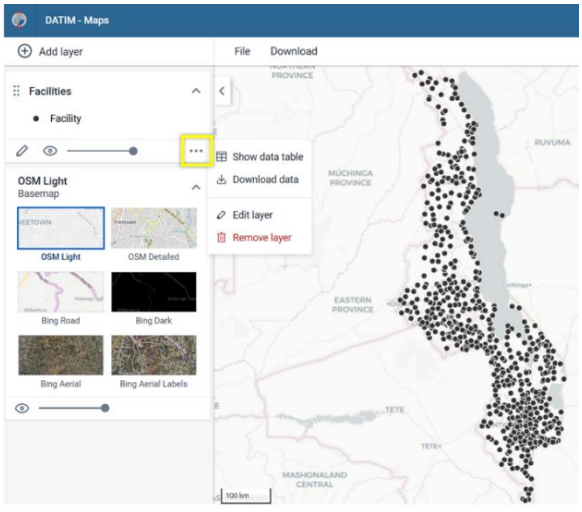
S/GAC's PRIME Team periodically develops a report that identifies health infrastructure which lack geographic coordinates in DATIM. PRIME sends that report to DATIM Site Administrators who should work to update DATIM with complete and accurate coordinate data.

3.2 Obtaining and working with facility latitude and longitude from DATIM

Country teams can obtain the facility-level spatial data in three ways. First, a call to the DATIM API will prompt a facility list with latitude and longitude to download. This option is the best way to obtain coordinate data in a table. For example, DATIM users with the appropriate permissions can download facility coordinates for Malawi, contained in a csv table, with the following query:

https://datim.org/api/organisationUnits.csv?filter=ancestors.name:eq:**Malawi**&filter=organisationUnitGroups.name:eq:Facility&fields=id,name,geometry

The permission to run these queries should be enabled for users as part of the credentials associated with your DATIM login. Second, the Map App interface includes a download data function. Within the Map App, users can load facilities to the map and navigate to the "Download data" option as shown.



Map App data download as GeoJSON, which is supported by most GIS software and some programing languages, such as QGIS, ArcGIS Desktop, R, and Python. Finally, country team members can obtain a current site list with geographic coordinates by filling out a helpdesk ticket with DATIM Support[2].

[2] DATIM Support Help Center, https://datim.zendesk.com/hc/en-us

3.3 Prioritizing geographic coordinates for capture or improvement

Spatial analysis to support the efficient allocation of resources requires complete and accurate coordinate data. Four filters can help identify which geographic coordinates a DATIM Site Administrator should prioritize for collection or improvement. A prioritization strategy is especially important for countries where PEPFAR supports thousands of health facilities.

1. **PEPFAR-supported sites.** Some facilities in DATIM do not receive PEPFAR support. Country teams should prioritize the improvement of coordinates for sites where PEPFAR supports programs.
2. **Subnational unit prioritization.** For some service coverage analysis, the spatial distribution of facilities where a service is available is important, whether PEPFAR supports the service in that location or not. Country teams might prioritize geographic coordinate quality and completeness in areas the program has prioritized for program focus, for example Scale-up Saturation and Scale-up Aggressive areas.
3. **Missing coordinates** are a priority. DATIM Site Administrators should attempt to assign geographic coordinates when adding new facilities to DATIM.
4. **Identified error is a flag for attention**. Coordinates that cause a facility to appear in the wrong administrative unit (e.g. district) require attention and possible correction.
5. **Duplicate coordinates** may be present. Unique, precise coordinates are much preferred.

3.4 Capturing geographic coordinates

Site visits are an opportunity to improve geographic coordinate data. The best coordinate data typically comes from direct capture with a GPS enabled device during a site visit. When geographic coordinates for a facility are missing or of insufficient quality, the country team or an implementing partner should capture and provide those coordinates to DATIM Site Administrators.

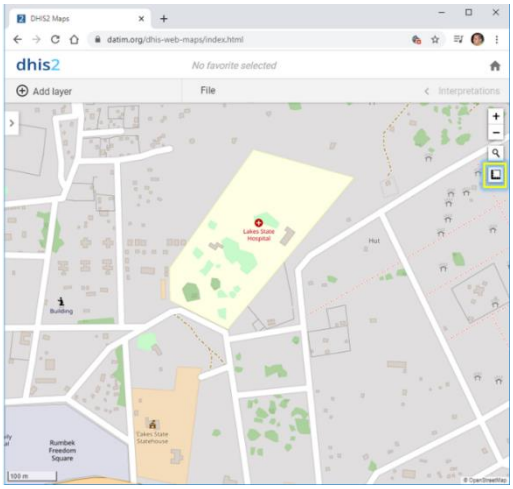
Technical aspects of geographic coordinate acquisition and maintenance are well-covered in WHO's Master Facility List Resource Package.[3] Consult chapter 6, "Geographic Coordinates in the MFL", for general instructions on primary data collection, such as integrating geographic coordinate capture into site visits. When collecting coordinates, use a current facility list snapshot from DATIM as the basis for data collection. This report should include the complete subnational unit hierarchy for each piece of health infrastructure and corresponding DATIM unique identifiers (uid). The linkage of new coordinate data to the uid is essential for accurate transfer into DATIM as well as other GIS-related applications.

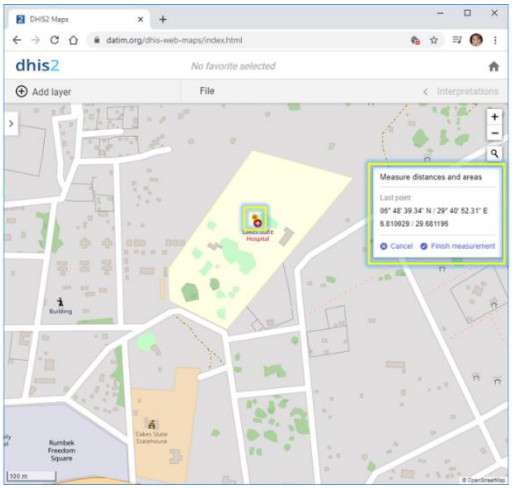
3.5 Other sources of geographic coordinates

It is not always necessary to conduct GPS survey at sites to fulfill the geographic coordinate requirement. What follows is a list of sources for geographic coordinates other than an MFL that may be readily at hand. The list sequences sources in the order that is likely to be most convenient and not necessarily in the order of the highest data quality:

[3] WHO. Master Facility List Resource Package: guidance for countries wanting to strengthen their Master Facility List. Geneva: World Health Organization; 2019. https://www.who.int/publications/i/item/-9789241513302

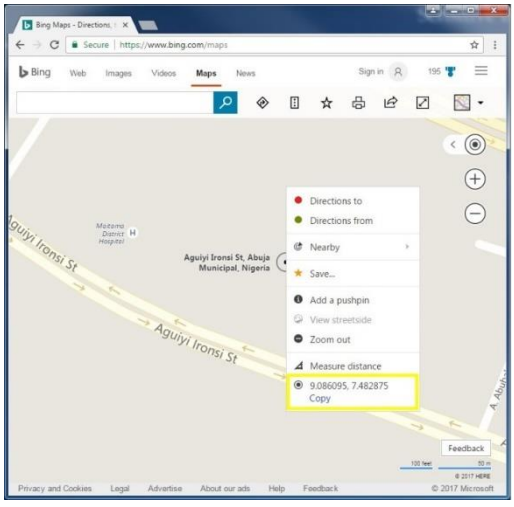
DATIM: The DATIM Map App measurement tools can help users identify coordinates for sites that lack coordinates and can help improve the precision of existing coordinates. If you can find the facility on the map using the map navigation tools (zoom, pan, etc.), you can right click on a map location to “Show latitude/longitude”.

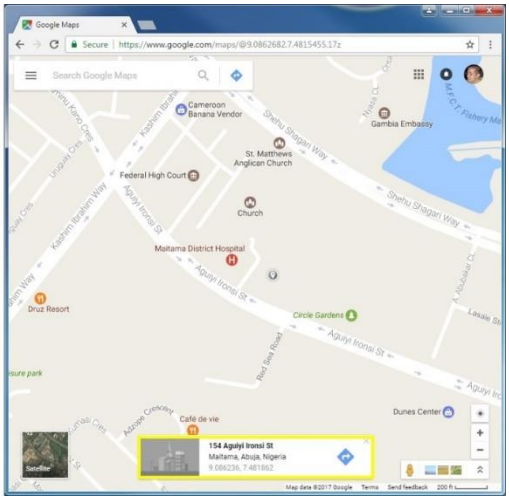




Shapefile: For every Organization Unit there is information on the location of at least some health facilities in an existing shapefile. These files may be found within the interagency space or among ministries of health, central statistical agencies, implementing partners, and common online spatial data repositories.

Web-based map platforms: Some web-based maps allow users to capture latitude and longitude coordinates that correspond to the location of the mouse pointer on the map. In Google Maps, users can left click on the map. Latitude and longitude for that point appears at the bottom of the map. The same information is available from Bing Maps with a right click. If an implementing partner can identify the location of a supported health facility on a map, these tools can identify coordinates for sites that lack coordinates or help improve the precision of existing coordinates. When using this technique, users are encouraged to zoom in to the exact location of the site before capturing coordinates.



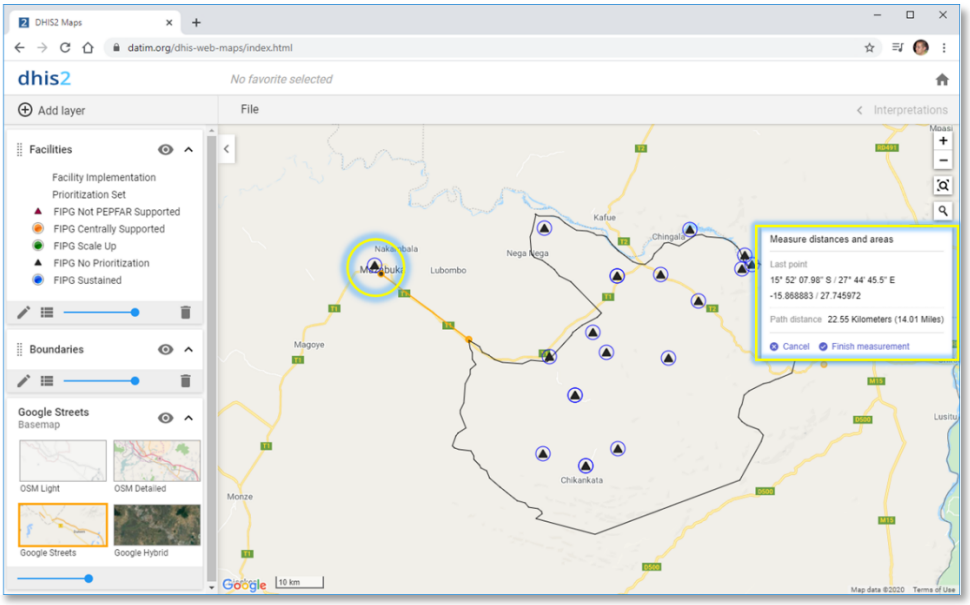


3.6 Quality checks for geographic coordinates

The quality of spatial analysis depends on the quality of spatial data. There are several common problems you can identify through a review of spatial data in spreadsheet format.

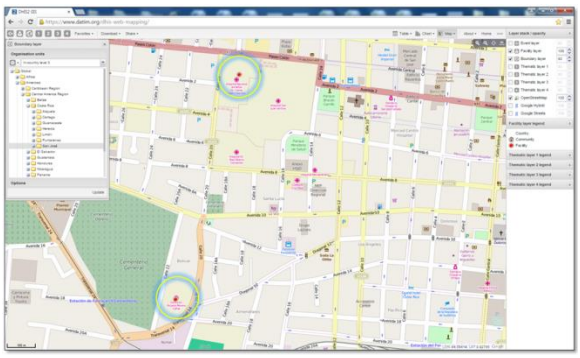
1. Transposed latitude and longitude usually cause points to appear far outside of the country in which they should appear. Confirm that the coordinate pairs have latitude in the latitude field and longitude in the longitude field.
2. Missing "-" signs also typically introduce major error. This sign indicates latitudes below the equator or longitudes west of the Prime Meridian.
3. Coordinates out of range, in the wrong format, or without decimals are apparent upon casual inspection.
    a. Range: Latitude measurements range from 0° to (+/–) 90°. Longitude measurements range from 0° to (+/–) 180°.
    b. The PEPFAR coordinate standard looks like this: 40.4463, -120.9821. Coordinate formats that look like the following require conversion: 26' 46" N 120° 58' 56" W; 40° 26.767' N 120° 58.933' W; 40N 630084 4833438
    c. Coordinates that have no values to the right of the decimal place have limited utility and are not adequate.

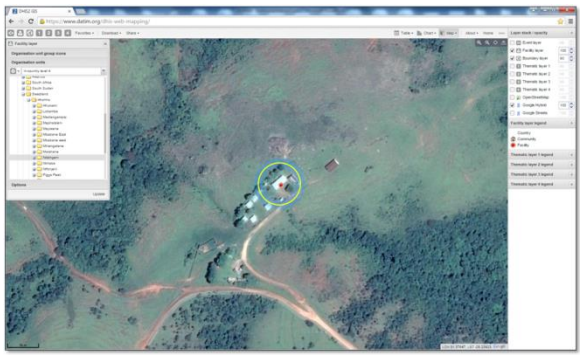
Some geographic coordinate errors are obvious, but some might not be apparent until visualized on a map. An easy means of reviewing coordinates is to see if the coordinates appear inside the administrative unit in which it should. The DATIM map app can easily visualize admin unit boundaries and associated facility coordinates. The screen capture below shows a site that is 22 kilometers outside of the district where it appears in the geographic hierarchy. The geographic coordinates could be wrong. It's also possible that the coordinates are accurate, but the site actually belongs in a different district in DATIM. DATIM Site Administrators can change the coordinates. Moving a facility to the correct place in the DATIM hierarchy requires assistance from HelpDesk.



Health facility outside of district

Another way to check geographic coordinate accuracy, shown below, is to visualize the point against the OpenStreetMap (OSM) or Google Hybrid layers in DATIM to see if the other data corroborate the geographic coordinate. For example, the left image below of San Jose, Costa Rica shows two health facilities from the DATIM database that appear directly next to health facilities in the OSM database. The image on the right shows the same level of accuracy for a facility in Eswatini.





3.7. Point standards

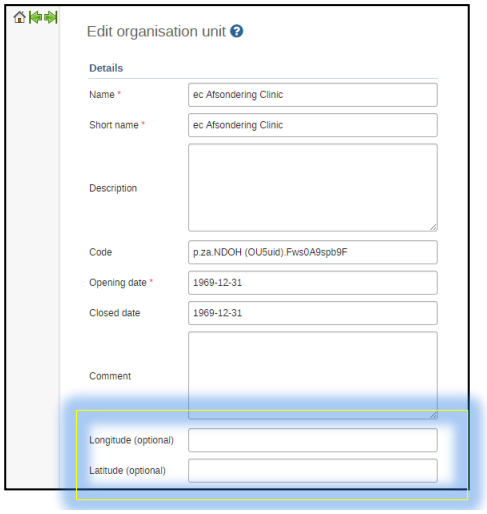
DATIM enforces the following standards on geographic data: coordinates must be recorded as geographic latitude/longitude in decimal degrees (e.g. -25.3443, 131.0344) using the WGS84 datum/reference ellipsoid (see below). We recommend recording locations to a precision of no less than four decimal places. Coordinates with values of only two decimal places could indicate a location more than one kilometer away from the intended location. Four decimal places are also the typical accuracy of a handheld GPS unit.

decimal places	degrees	N/S or E/W at equator
0	1.0	111.32 km
1	0.1	11.132 km
2	0.01	1.1132 km
3	0.001	111.32 m
4	0.0001	11.132 m
5	0.00001	1.1132 m

Decimal places and precision

If you have coordinates in an alternative format and require assistance to convert them to decimal degrees, please contact DATIM Support.

3.8. Updating geographic coordinates in DATIM

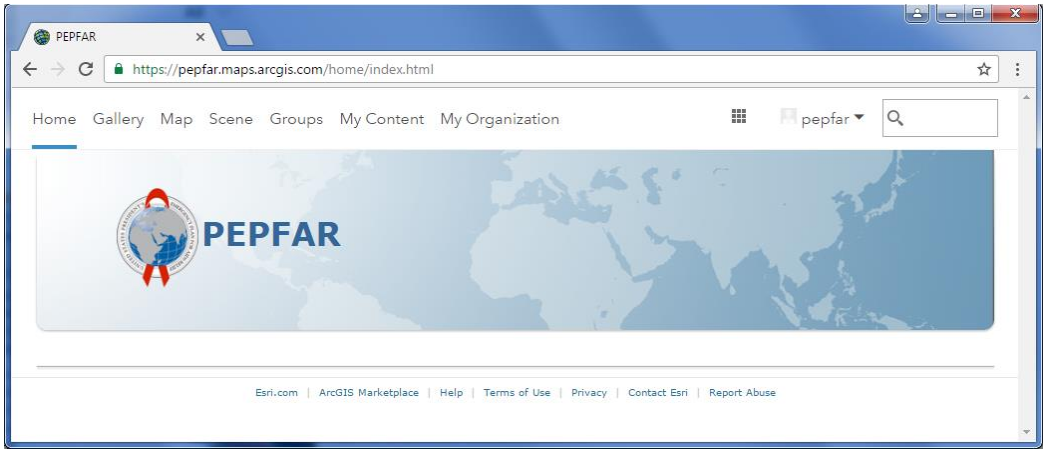


Latitude and longitude as attributes of an organization

The organization unit profile is where the DATIM Site Administrator will add coordinates for new clinical sites or update coordinates when improved data become available. For bulk import of large numbers of site latitude and longitude, contact DATIM Support.

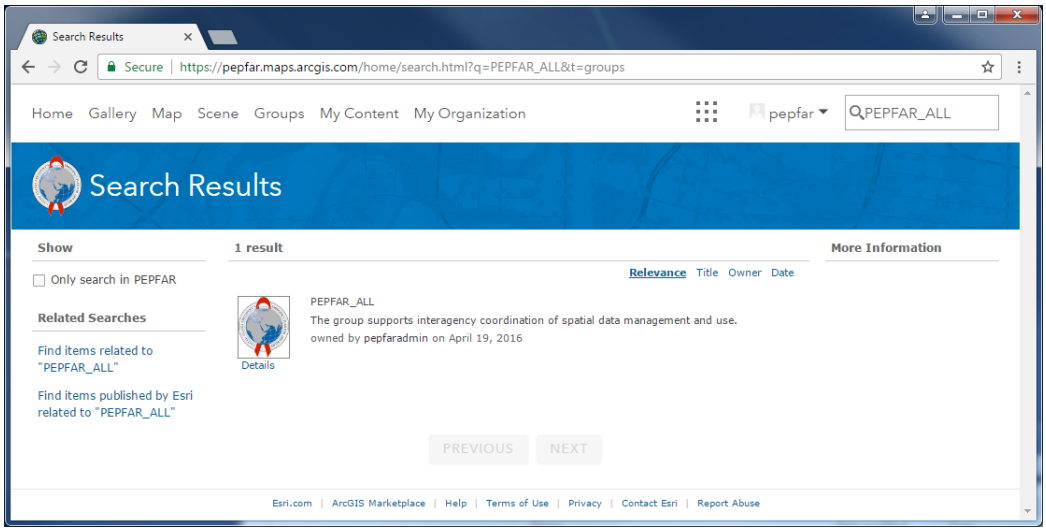
4. Polygons

PEPFAR maintains polygonal spatial data that represent the geography of PEPFAR program planning. Polygons may represent a standard administrative hierarchy or specialized geography demarcated by a ministry of health. Subnational unit hierarchies do not need to be comprehensive. For example, if a country hierarchy includes provinces, states, and districts, the hierarchy might omit the state level if the level is not relevant to health planning. However, hierarchies should always align with the standard set by the relevant authority in the partner government.



PEPFAR stores polygons with ArcGIS Online using an organizational account.[4] The PEPFAR Organizational Account is the system of record for polygonal spatial data used in web mapping applications such as DATIM, Panorama, and Panorama Spotlight.[5] PEPFAR's ArcGIS Online Account is a digital service provided to support interagency collaboration for the development and maintenance of PEPFAR's spatial data. The Organizational Account shares data to an ArcGIS Online group, PEPFAR_ALL. The PEPFAR community can gain direct access to PEPFAR_ALL by logging into an agency Esri organizational account, searching for the PEPFAR_ALL Group and requesting to join the group.

Polygon origins vary but are typically from common public repositories. To facilitate data joins, file attribute tables include the DATIM uid and name for the file's DATIM level and each parent level in the hierarchy. At the geographic level for which data have been cleared for release on PEPFAR Spotlight, neither the polygons nor the data contained within the attribute tables are programmatically sensitive, either on their own or together. Therefore, most polygons are open to the public for the geographic levels at which PEPFAR data have been cleared for publication to PEPFAR Spotlight.



PEPFAR's shapefile naming convention in ArcGIS online is as follows:

- Country
- DATIM level
- Administrative unit name
- File date

An underscore separates all name elements: Botswana_5_HealthDistricts_2022_Nov. The use of this naming convention allows easy extraction of data from ArcGIS Online for use in updates to Panorama and Panorama Spotlight.

[4] PEPFAR ArcGIS Online, https://pepfar.maps.arcgis.com/home/index.html
[5] Panorama Spotlight, https://data.pepfar.gov/

4.1 Polygons: Roles and responsibilities

4.1.1 Country teams

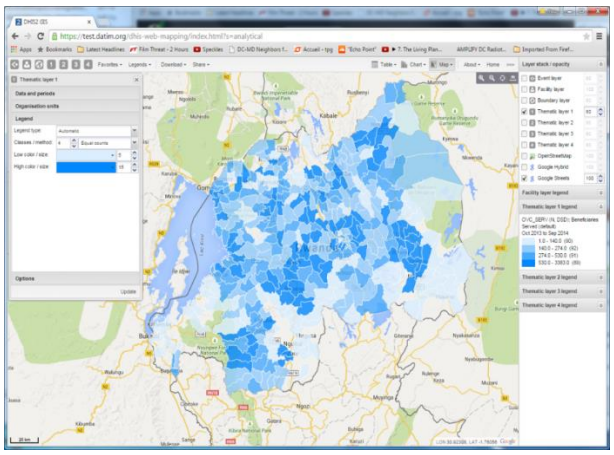
Country teams should periodically review maps in DATIM and Panorama for completeness and accuracy. In some cases, entire levels of the organization unit hierarchy are missing in the DATIM and Spotlight maps due to a lack of available matching map data (e.g. a ward shapefile). Missing map data can affect analysis. Country teams should proactively seek to fill map gaps through outreach to partner governments and local GIS communities of practice. Central statistical agencies or ministries of health are likely sources of authoritative shapefiles, whether they represent administrative boundaries or health districts. In some country contexts there are national mapping agencies that have the responsibility for developing authoritative digital maps.

Boundaries can change. When partner governments change administrative units there may be structural implications for the organizational unit hierarchy in DATIM (e.g. one district splits into two). Deciding how to manage hierarchy changes in DATIM requires a broad conversation within the team, in consultation with DATIM Support. DATIM Support will work with the country team to update maps once any needed adjustments are made to the organization unit hierarchy.

Country Teams can send digital polygons (shapefiles, etc.) that accurately represent new administrative boundaries to DATIM Support. If able, the team should add a DATIM uid column to the attribute table and link the uid code to its respective polygon.

4.1.2 PRIME

PRIME is responsible for periodically checking which geographic entities lack geometry and working with DATIM Site Administrators to ensure that DATIM and Panorama maps are up to date. PRIME also confirms with The U.S. Department of State's Office of the Geographer and Global Issues that PEPFAR spatial data conform to official policy on representations of international boundaries.[6]



Example of polygon data in DATIM

[6] U.S. Department of State GeoNode. Large Scale International Boundaries (LSIB). http://geonode.state.gov/

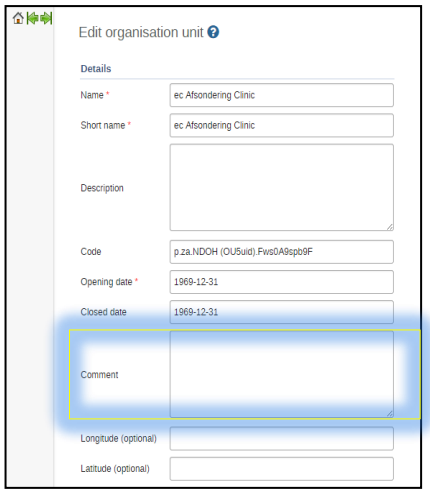
4.2 Importing polygons into DATIM

DATIM Support adapts shapefiles for use in DATIM and loads polygons into the DATIM and Panorama environments.

5. Spatial metadata

Data sets should have accompanying metadata. At a minimum, metadata should include the following elements: Title, Description, Tags, Last Update, Publisher, Contact Name, Unique identifier, Public Access Level, and License. A description could include a wide variety of information such as the file's source, the edit history of the file, or known discrepancies in the alignment of the file and the DATIM hierarchy.

Spatial data stored in ArcGIS Online should have a complete metadata profile for both stored shapefiles and feature layers. DATIM currently has a limited ability to record metadata specific to spatial data. Even so, in keeping with good data management practice, information such as the geographic coordinate source, year of collection and method of collection (e.g. GPS) could be documented in the comment section of the organization unit profile.



Comment section of an organization unit profile

5.1 Geographic exactness

DATIM Site Administrators can also use the codes below (1 or 2) to indicate coordinate precision. In the early stages of geographic data development, it is acceptable to use approximate geographic coordinates for a clinical site. Unless a team has surveyed the clinical site with a GPS device or extracted the exact location of the site from high resolution satellite imagery, use the Geographic Exactness code 2 (approximate). Over time, SIMS and other site visits will provide opportunities to replace approximate coordinates with exact, GPS-derived coordinates.

Code	name
1	Exact
2	Approximate

5.2 Geographic class

Additional optional codes allow you to indicate the geographic class, or scale, of your coordinates. This data element is not required. However, the code works with the precision code to indicate what a pair of geographic coordinates represents. For instance, clinical sites are understood to be “brick-and-mortar” structures but the coordinates used to describe them may, in some cases, only be as precise as the populated place in which the clinical site exists. This would result in a precision of 2 (approximate) and a class of 2 (populated place). A GPS survey of the clinical site would provide a precision of 1 (exact) and a class of 3 (structure).

Code	name
1	Administrative Region
2	Populated Place
3	Structure
4	Other Topographical Feature

6. Mobile clinics

PEPFAR represents mobile clinics with points. However, the geography of these sites is not static and may encompass a large area. For locating mobile clinics, the geographic exactness code should be 2, "approximate". The latitude and longitude can represent a facility where the mobile clinic is based or the rough center point of the populated place or low-level administrative unit the mobile clinic serves. The geographic class code should then be either 2, "populated place" or 3, "administrative region". Unless the administrative region is quite granular and a good fit for the mobile clinic's range, we would recommend using the populated place. If a clinic’s mobile nature is not apparent by its name, it’s possible to note this in the comment section of the organization unit profile.

7. Datum

Points and polygons are used in relation to a datum, which is a mathematical model for representing the shape of the earth. While this is a technical point, it’s important. Earth’s shape is irregular. Some datums are better suited to specific areas of the globe. The WGS84 datum provides a good average for use across the globe and is what DATIM and Panorama use. If we reference latitude and longitude in the wrong datum, coordinates could be off by large distances when mapped. Please confirm that GPS-enabled devices are set to WGS84 before collecting coordinates. Also confirm the datum when receiving spatial data from implementing partners. Failure to track this information can render spatial data useless.

8. Sensitive spatial data

In general, PEPFAR site lists are not inherently sensitive. However, linked spatial and programmatic data can trigger a need to pay heightened attention to privacy, confidentiality, and security concerns. For example, absent rigorous masking techniques, maps of facility-level key population data are not releasable to the public.

PEPFAR site lists should reflect the true name of the facility and not programmatic references, especially concerning key populations (e.g. Sex Worker Friendly Clinic #8). If a military site is accidentally introduced into DATIM, DATIM Site Administrators should delete any associated geographic coordinates immediately and alert DATIM Support Helpdesk.

DATIM site lists with latitude and longitude are accessible to PEPFAR USG Staff to support location-based analytics and inform PEPFAR programming. These lists may include sensitive information that should be redacted in analytic contexts. To mitigate risk and prevent the inadvertent distribution of sensitive data, DATIM Site Administrators should periodically coordinate with country teams and partners to review PEPFAR site lists for potentially sensitive information, such as military bases or problematic key population-related programmatic references.

DATIM Site Administrators have the authority to change facility names and should work with implementing partners to confirm accurate site names and maintain DATIM accordingly. DATIM Site Administrators for each OU, and their contact information, can be found in the PRIME contact list on PEPFAR SharePoint. Country teams should submit a DATIM Support Helpdesk ticket requesting guidance on how to address discovered military sites. If your team has concerns about the potential sensitivity of spatial data in DATIM, please contact DATIM Support.

PEPFAR attenuates risks associated with potentially sensitive data through geographic aggregation where appropriate. [7, 8] In addition to geographic aggregation, many cartographic techniques for visualizing sensitive spatial data are commonly used, such as geographical masks (e.g. donut) or spatial obfuscation.

The most granular data used for routine planning and monitoring are at the site level, which is either a clinical site or a low-level administrative unit (e.g. a district). Data may be anonymized to allow publication at fine levels of granularity. Additionally, the program may use cell suppression for small values to reduce risks associated with disclosure of these data.

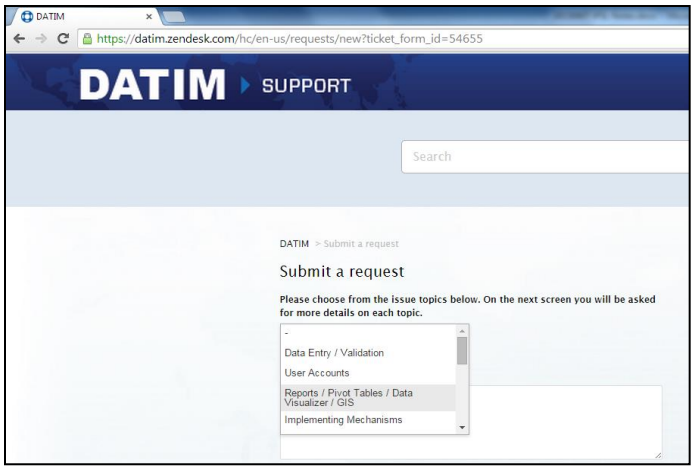
---

[7] National Institute of Standards and Technology. Guide to Protecting the Confidentiality of Personally Identifiable Information (PII). Special Publication (NIST SP) - 800-122, https://www.nist.gov/publications/guide-protecting-confidentiality-personally-identifiable-information-pii

[8] PEPFAR Data Governance Guidance, https://datim.zendesk.com/hc/en-us/articles/360061233211-PEPFAR-Data-Governance-Guidance

9. GIS and DATIM Support

DATIM Support can respond to a wide array of GIS questions as outlined in this document. To request central support, submit a DATIM Support Helpdesk ticket.



DATIM Support screen

10. Map clearance and publication

Maps should undergo a public affairs clearance process before circulating outside of the U.S. Government. In addition to a thorough review of map content, include this checklist:

- ✓ Date of production
- ✓ Data source
- ✓ Contact information
- ✓ Consult ArcGIS Online and use LSIB-aligned shapefiles when possible.
- ✓ Include disclaimer, “Names and boundary representation are not necessarily authoritative.”
- ✓ When mapping health facilities, note the number of facilities for which there are coordinates. For example, “Map shows X of Y PEPFAR supported health facilities that reported HTC_TST_POSITIVE.”
- ✓ When mapping health facilities, include disclaimer, “Some health facilities may appear in places other than their true location.”
- ✓ Consider factors that would cause the map to be sensitive.

11. National spatial data infrastructure alignment

Maps fill an essential coordination function in health planning. Alignment with partner government reference geography supports data exchange and joint planning among all parties that adopt that standard. DATIM geographic hierarchies align with the province, district, and other nested geographic structures established by partner governments. The facility list in DATIM likewise aims to be an accurate, aligned reflection of health infrastructure in a country.

PEPFAR-supported countries generally have a government GIS unit or a dedicated staff member within a National Mapping Agency or Bureau of Statistics who has a mandate for the management of spatial data. Institutional linkages should be made and maintained to ensure that PEPFAR and partner government spatial data infrastructure are well aligned. PEPFAR supports national spatial data initiatives framed by policies that seek to coordinate the collection, storage, maintenance, and sharing of spatially referenced data.

11.1 Polygons

When countries alter subnational boundaries (e.g. district splitting), PEPFAR first reflects the change in DATIM's organization unit hierarchy. Once DATIM automatically assigns the new units a uid, it's then possible to link geometry to PEPFAR data reported to these units and load updated polygons to DATIM and Panorama.

11.2 Facility latitude and longitude

Geographic coordinates will likely be present in an MFL or another partner government-managed official list. Geographic coordinates are one component a country's MFL and country teams should work to align DATIM and MFL coordinates. The best available guidance on this topic is WHO's Master Facility List Resource Package, which addresses MFL governance as well as technical aspects of geographic coordinate acquisition and maintenance.[9] When accurate infrastructure location data is generated through PEPFAR, the resulting datasets should be forwarded to appropriate contacts in national mapping agencies or other appropriate ministries.

12. Data formats

Spatial data generated through PEPFAR should be stored in a machine readable, non-proprietary, open data format (e.g. shp, GeoJSON, and kml).

13. The PEPFAR GIS Community of Practice

To support a GIS community of practice, ICPI and S/GAC have curated relevant geospatial resources in a wiki. If you have questions, would like to add resources to the wiki, or engage with the GIS team, please tag us on GitHub as @ICPI/GIS or send a message to the PEPFAR GIS listserv. The PEPFAR GIS listserv is open to all GIS practitioners, cartographers, and map enthusiasts with a .mil or .gov address. The group supports interagency sharing of information, resources, and experience related to the use of GIS for HIV epidemic control. To sign up, send a message to PEPFAR-GIS-subscribe-request@listserv.gsa.gov.

[9] WHO. Master Facility List Resource Package: guidance for countries wanting to strengthen their Master Facility List. Geneva: World Health Organization; 2019. https://www.who.int/publications/i/item9/-9789241513302